Simultaneous modeling of initial conditions and time heterogeneity in dynamic networks: An application to Foreign Direct Investments

JOHAN KOSKINEN

Social Statistics Discipline Area, University of Manchester, Manchester M13 9PL, UK
(e-mail: johan.koskinen@manchester.ac.uk*)*

ALBERTO CAIMO and ALESSANDRO LOMI

Faculty of Economics, University of Lugano, Switzerland

Abstract

In dynamic networks, the presence of ties are subject both to endogenous network dependencies and spatial dependencies. Current statistical models for change over time are typically defined relative to some initial condition, thus skirting the issue of where the first network came from. Additionally, while these longitudinal network models may explain the dynamics of change in the network over time, they do not explain the change in those dynamics. We propose an extension to the longitudinal exponential random graph model that allows for simultaneous inference of the changes over time and the initial conditions, as well as relaxing assumptions of time-homogeneity. Estimation draws on recent Bayesian approaches for cross-sectional exponential random graph models and Bayesian hierarchical models. This is developed in the context of foreign direct investment relations in the global electricity industry in 1995–2003. International investment relations are known to be affected by factors related to: (i) the initial conditions determined by the geographical locations; (ii) time-dependent fluctuations in the global intensity of investment flows; and (iii) endogenous network dependencies. We rely on the well-known gravity model used in research on international trade to represent how spatial embedding and endogenous network dependencies jointly shape the dynamics of investment relations.

Keywords: *dynamic networks, longitudinal exponential random graph model, foreign direct investment network, gravity model, spatial processes, exchange sampler, world trade network topology, Bayesian data augmentation, continuous-time Markov chains*

1 Introduction

We consider longitudinal models for network ties with a focus on a class of generative models defined for continuous time. We are not concerned with discrete-time models such as that proposed by Robins & Pattison (2001). By couching the process of tie-change in terms of an embedded chain, where the decisions of actors incrementally changes the entire network, Snijders' stochastic actor-oriented model (SAOM) (Snijders, 2001) relaxed the independence assumptions of the earlier continuous-time models (Holland & Leinhardt, 1977; Wasserman, 1980). The exponential random graph model (ERGM) was developed explicitly with the aim of modeling dependencies among tie-variables and the model has also been extended to modeling tie-change in (continuous) time (for a description see e.g., Snijders, 2006; Snijders

& Koskinen, 2013; Koskinen & Lomi, 2013). When the continuous-time model is only observed in discrete time $t_0, t_1, \ldots, t_{M-1}$, $t_m \in \mathbb{R}^+$, it is difficult to draw inference without assuming that the process is homogenous in time. Because these are models for *change*, conditioning on the first observation at t_0 greatly facilitates inference. We propose an approach for making use of the information contained in the first observation, rather than merely using it to assess the change from one observation to the next. We couple this with proposing an approach for addressing time-heterogeneity in a parametric framework.

We illustrate the proposed approach by elaborating on the analysis of foreign direct investment (FDI) in the international electricity industry of Koskinen & Lomi (2013). FDI are international capital flows determined by investment decisions taken by a company in one country (parent company) to acquire control over a company in a different country (target company). FDI decisions at the company level establish a specific kind of network between countries whereby capital resource flows into the country of the target company in exchange for ownership and voting rights that flow in the opposite direction toward the country of the parent company. The dyadic exchanges underlying FDI cumulate into a network of dependence relations between countries in the world economy.

The motivations and strategies behind FDI decisions are typically contingent on a variety of highly heterogeneous firm and industry-specific factors (Redding, 2011). By concentrating on FDI in the international electricity industry (a single-industry design) we reduce the large set of possible sources of unobserved heterogeneity. Koskinen & Lomi (2013), furthermore, selected the electricity industry for study because FDI played a major role in the globalization of the electricity industry. During the observation period (1994–2004) the electricity industry underwent rapid globalization and a rapid transition from a set of disconnected national monopolies to a truly global industry. Because we are interested in relations between countries emerging from an investment relation linking existing companies, we do not deal with greenfield investments—a form of FDI where a parent company starts a new venture in a foreign country.

To capture the spatial embedding of ties in an ERGM, Daraganova *et al.* (2012) specified a distance interaction function that Koskinen & Lomi (2013) then used to specify a model for FDI relations that was in turn an adaption of the well-known gravity model of trade (Anderson, 1979; Bergstrand, 1985). Well established in the study of international trade (Anderson, 2011; Anderson & van Wincoop, 2003), during the last decade the gravity model has been successfully applied also to FDI relations between countries (Bevan & Estrin, 2004; Chakrabarti, 2003). FDI originate from individual investment decisions that are typically non-repeatable events. Hence, it would seem that a model derived for explaining trade flows cannot be directly applied to FDI decisions. We are not modeling individual investment decisions (taken at the company level), but aggregate relations that these decisions involve (at the country level). In other words, we assume that ties between countries are states rather than events. As such, relations between countries involve some degree of inertia and are not modified immediately by additional investment events—or their lack thereof.

The gravity model is useful for our purposes because it involves an explicitly dyadic formulation. In its basic specification, the gravity model relates bilateral

flows between two countries (Y_{ij}) to their size (V_i and V_j—as measured, for example by gross domestic product (GDP)) and their distance (D_{ij}). In its simplest form, the gravity equation specified and estimated in empirical models is therefore: $Y_{ij} = V_i V_j / D_{ij}$ where the effect of size is expected to be positive to signal trade opportunities and the effect of distance is expected to be negative to signal fixed trade costs.

When the focus is on the change in the underlying topography of the FDI network (Dueñas & Fagiolo, 2013; Squartini *et al.*, 2011a) the network ties underlying observed capital flows are of interest, not the capital flows themselves. The topography of the FDI network is represented in terms of binary, directed ties between countries. We let the tie-variable X_{ij} be one or zero depending on whether there is an FDI tie from country i to country j or not, respectively for $i, j \in \{1, \ldots, n\}$. The space of all adjacency matrices is denoted $\mathcal{X} = \{0, 1\}^{V^{(2)}}$. Considering the topography of the FDIs allows us to investigate the effect of distance while allowing us to draw on the elaborate results for dependence offered by ERGMs.

An exclusive focus on network structure may be motivated, in part, by recent results suggesting that the structure of ITN may be fully characterized in terms of their local topological properties (Squartini *et al.*, 2011a). More specifically, we focus on the binary architecture of the FDI networks because we are interested in documenting the emergence and the effect of local (dyadic and triadic) network substructures. As Fagiolo *et al.* (2009) clearly show in the context of international trade, weighted networks display a weaker clustering due to the presence of many low-intensity trade flows. Networks of FDI flows are likely to display similar properties that would be undesirable given the objectives of our study, namely to investigate how the nature of dependencies change over time.

2 The model

We define an ERGM *process* as the process on $\mathcal{X}$ for which the limiting distribution is an ERGM (Holland & Leinhardt, 1981; Frank & Strauss, 1986; Wasserman & Pattison, 1996; Snijders *et al.*, 2006) of the form

$$\Pr_\theta(X = x) = \exp\{g(x; \theta) - \zeta(\theta)\} \tag{1}$$

where $g(\cdot; \theta)$ is some function of the adjacency matrix x and the vector of parameters $\theta \in \Theta \subset \mathbb{R}^p$, and $\zeta(\theta) = \log \sum_{x \in \mathcal{X}} e^{g(x;\theta)}$ is a normalizing constant that ensures that the distribution sums to unity on $\mathcal{X}$ which is typically computationally intractable. As described in for example Snijders (2006), the ERGM process may be defined as a process $X(t)$, for a time-parameter $t \in T \subset \mathbb{R}^+$, that is a continuous-time Markov chain with intensity matrix Q, but that evolves through incremental changes to the network. This is the same basic principle as in Holland & Leinhardt (1977) and in the SAOM (Snijders, 2001) where the incremental changes are defined in terms of toggles of tie-variables. The tie-based ERGM process, or longitudinal ERGM (LERGM), does however differ from SAOM in some central assumptions as well as in some of the details of the modeling and estimation (Koskinen & Lomi, 2013; Snijders 2006; Snijders & Koskinen, 2013).

In order to define a toggle, for $x \in \mathcal{X}$, let x_{-ij} be the incomplete adjacency matrix $\{x_{uv} : (u, v) \in V^{(2)} / \{(i, j)\}\}$, which is equal to x from which the information about

the value of the tie variable x_{ij} is deleted. Define the operator $\Delta_{ij}^{\text{sign}}$, such that for x, $y = \Delta_{ij}^{\text{sign}} x$ has $y_{-ij} = x_{-ij}$ and $y_{ij} = 1$, $y_{ij} = 0$, or $y_{ij} = 1 - x_{ij}$ according to whether sign is equal to $+$, $-$, or $*$, respectively.

Thus $\Delta_{ij}^* x$ toggles the entry (i, j) of x and we define the neighborhood of a graph x, as $N(x) = \{y \in \mathcal{X} : y = \Delta_{ij}^* x,$ for some $(i, j) \in V^{(2)}\}$, and define the Markov chain in terms of a process that stays in the current state x for some time and then jumps to a state in the neighborhood of x or itself. In terms of Q this implies the strictly positive rates $q(x, y) = q_{ij}(x)$ for $y = \Delta_{ij}^* x$ in the neighborhood of x but $q(x, y) = 0$ for y that are not in $N(x) \cup \{x\}$. This Markov chain is an ERGM process if we define the rates as $q_{ij}(x) = \rho \Pr_\theta(X_{ij} = 1 - x_{ij}|X_{-ij} = x_{-ij})$, where

$$\Pr_\theta(X_{ij} = 1|X_{-ij} = x_{-ij}) = [1 + \exp\{g(\Delta_{ij}^- x; \theta) - g(\Delta_{ij}^+ x; \theta)\}]^{-1} \qquad (2)$$

is the conditional tie-probability of the ERGM. The interpretation is that if randomly chosen tie-variables are updated using the Gibbs sampler with probabilities (2), then the limiting distribution is Equation (1). In the context of FDI, the model explains why ties are present at different points in time as a function of the spatial embedding and resources of states while accounting for the fact that ties are not independent and that there were ties at a previous time-point. A tie-variable is thus both spatially embedded, embedded in a "social neighborhood" (Pattison & Robins, 2002), and embedded in time, being subject to the temporal dynamics emanating from past history.

3 Estimation and extension

While the ERGM is an exponential family distribution, the ERGM *process* is not and the standard estimation procedures for the ERGM cannot be used. However, *conditional* on the initial state $x(t_0)$ the model parameters may be estimated using similar estimation techniques to the ones used for the SAOM such as stochastic approximation (Snijders, 2001) or Bayesian data augmentation (Koskinen & Snijders, 2007). Extending these estimation schemes to the case of joint estimation for all of data is not trivial other than for trivial auxiliary models for $x(t_0)$. Before we introduce the modeling extensions we briefly describe conditional estimation for the process given $x(t_0)$. The reason being that the definition of a model for $x(t_0)$ and the time-varying parameters affect the extent to which estimation is tractable. Thus modeling and estimation considerations are developed in tandem.

3.1 Conditional estimation for the LERGM

The Bayesian data-augmentation scheme for the LERGM conditional on the initial state involves setting up a Markov chain Monte Carlo (MCMC) scheme that alternates between drawing from the conditional posterior distribution of the parameters and sampling paths that link the data that are observed in discrete time. The augmentation of data is necessary as the likelihood based only on observed data is given by elements of a transition matrix $P(t) = e^{tQ}$ that is intractable.

For data $x(t_0), \ldots, x(t_{M-1})$ observed at observation moments $t_0 < t_1 < \cdots < t_{M-1}$, a sample path is a sequence of graphs $v \in \prod_{m=1}^{M-1} \mathcal{X}(x(t_{m-1}), x(t_m))$, for $\mathcal{X}(x, y) = \{(v_0, v_1, \ldots, v_R) \in \mathcal{X}^R : v_0 = x, v_R = y, v_r \in N(v_{r-1}) \cup \{v_{r-1}\}\}$. The variate $v \in$

$\mathcal{X}(x(t_{m-1}), x(t_m))$ is a sample path that is constrained to start in $x(t_{m-1})$ and end in $x(t_m)$, here the observed states we know the path must connect. Thus, while the data likelihood is intractable, the augmented data likelihood given a sample path v is given by the expression

$$L(\theta, \rho; v, x(t_0), \ldots, x(t_{M-1})) \propto \vartheta_\rho(R) \prod_r T_\theta(v_r; v_{r-1}), \qquad (3)$$

where R is the length of the sequence $v = (v_0, \ldots, v_R)$, the transition probabilities $T_\theta(v_r; v_{r-1})$ are given by Equation (2), and $\vartheta_\rho(R) = e^{-\rho/[n(n-1)](t_{M-1}-t_0)}/R![\rho(t_{M-1} - t_0)]^R$ relates the rate parameter to the number of steps in the path. We may sample the posterior variates, v, θ, and ρ by recognizing that the full conditional posteriors are proportional to Equation (3), multiplied by their priors in the case of θ and ρ.

The posterior variates are updated in three blocks: (a) updating of sample paths v, (b) updating of the LERGM parameters of Equation (2), and (c) updating of the rate parameters. Step (a) is a Metropolis updating step, where a move is proposed to v^*, drawn from the proposal distribution $D(v^*|v)$ conditional on the current state v. Following Koskinen & Snijders (2007) and Snijders *et al.* (2010), the proposed move is either constructed by lengthening v by two self-canceling moves Δ_{ij}^*, setting $v^* = (v_0, v_1, \ldots, v_s, \Delta_{ij}^* v_s, \Delta_{ij}^* v_{s+1}, \ldots, \Delta_{ij}^* v_r, \Delta_{ij}^* \Delta_{ij}^* v_r = v_r, \ldots, v_R)$; removing two canceling moves; or inserting or deleting an extra step that does not toggle the previous graph (a so called diagonal move). The update is accepted and $v := v^*$ with probability $\min\{1, H\}$, where the Hastings ratio

$$H = \frac{\vartheta_\rho(R^*) \prod_{r=1}^{R^*} T_\theta(v_r^*; v_{r-1}^*)}{\vartheta_\rho(R) \prod_{r=1}^{R} T_\theta(v_r; v_{r-1})} \frac{D(v|v^*)}{D(v^*|v)}.$$

Updating step (b) is similarly carried out with a simple Metropolis updating step, by proposing a move $\theta^* \sim F(\theta^*|\theta)$, from the current state θ, and then accepting the move using the Hastings ratio

$$H = \frac{\prod_r T_{\theta^*}(v_r; v_{r-1})}{\prod_r T_\theta(v_r; v_{r-1})} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{F(\theta|\theta^*)}{F(\theta^*|\theta)}.$$

As the transition probabilities for the augmented path are based on Equation (2) the acceptance probability is fully tractable. The rate parameter, given everything else, is drawn using a Gibbs updating step from the full conditional posterior gamma$(\rho; R + \alpha_0, [n(n-1)](t_{M-1} - t_0) + \beta_0)$. This differs from the implementation in Koskinen & Lomi (2013) where the parameterization of ρ is similar to that of Snijders & Koskinen (2012).

3.2 Estimating initial conditions

If we want to relax the conditioning on $x_0 = x(t_0)$, assuming that we want to leverage the fact that x_0 may contribute structural information, we need to define a model for x_0. This will allow us to base inference on the joint model

$$p(x_0, x(t_1), \ldots, x(t_{M-1})|\theta, \rho, \psi) = p(x(t_1), \ldots, x(t_{M-1})|\theta, \rho, x_0)p(x_0|\psi)$$

for all of data. We could chose as our model for x_0 a simple, analytically tractable model, in which case the inferential task would reduce to estimating an auxiliary model. However, it seems arbitrary to specify one model for the initial observation

and another completely unrelated model for the subsequent observations. We propose instead to assume that x_0 follows an ERGM defined by the same effects as in the LERGM with the corresponding parameters denoted by ψ. The fact that x_0 and the consecutive observations contribute the same type of structural information through $g(\cdot\,;\cdot)$ facilitates interpretation. We may address the question of to what extent structure is already present at t_0 and to what extent the structural "biases" are limited to the following network evolution and we may phrase this question in terms of the network configurations of the models. In particular, as the ERGM is the limiting distribution of the LERGM we may interpret x_0 as the outcome of a LERGM starting in $x(-s)$ for s large (note that this is not the same as setting $x(-s) = 0$ for arbitrary and fixed $s > 0$). Formally, with $\psi = \theta$, the formulation implies a process in equilibrium.

While we may update ρ as before, updating both θ and ψ in step (b) involves drawing from the full conditional posterior of (θ, ψ) given (x_0, v, ρ). This distribution is proportional to

$$\pi(\theta, \psi)p_\psi(x_0) \prod_r T_\theta(v_r; v_{r-1})$$

which in addition to the part stemming from Equation (3) is a function of the ERGM likelihood $p_\psi(x_0) = \exp(g(x_0; \psi) - \zeta(\psi))$ involving the intractable $\zeta(\psi)$. In the following, for the purposes of a convenient notation we consider the path v as fully observed, treating $T_\theta(v_r; v_{r-1})$ as the observed data likelihood and neglecting the updating step (a). We demonstrate how estimation is a straightforward extension for the case where the parameters are distinct. As discussed above this implies that the initial state distribution and rest are estimated separately. If parameters ψ and θ are constrained to be the same, estimation for the ERGM borrows power from the LERGM but, as we shall see, this complicates estimation. A solution is to stochastically couple the parameters through an additional layer of parameters.

3.2.1 Modified updating step for model parameters

If ψ and θ are distinct parameters, step (b) can be done in two blocks: update θ as before, and then update ψ. Given a completely augmented sample path, we can update ψ using the principle of the "exchange sampler" (Murray *et al.*, 2006). This involves drawing replicate data y and parameters ξ according to

   i. draw $\xi \sim h(\xi|\psi)$,
   ii. generate one network y from an augmented ERGM likelihood $p_\xi(y)$

where $p_\xi(y) = \exp\{g(x; \xi) - \zeta(\xi)\}$, and then set $\psi = \xi$ with probability $\min\{1, H\}$, for

$$H = \frac{p_\xi(x_0) \prod_{r=1}^{R} T_\theta(v_r; v_{r-1})}{p_\psi(x_0) \prod_{r=1}^{R} T_\theta(v_r; v_{r-1})} \frac{h(\psi|\xi)p_\psi(y)}{h(\xi|\psi)p_\xi(y)} = \frac{p_\xi(x_0)}{p_\psi(x_0)} \frac{h(\psi|\xi)p_\psi(y)}{h(\xi|\psi)p_\xi(y)}.$$

The update of ψ in (b) thus reduces to an approximate exchange sampler update (Caimo & Friel, 2011), and (b) may be understood as sampling from the following augmented distribution:

$$\pi(\xi, y, \theta, \psi|x_0, v_L, \ldots, v_1, \rho) \propto p_\psi(x_0)p(v_t, \ldots, v_{t-1}|\theta, x_0, \rho)\pi(\theta, \psi, \rho)h(\xi|\psi)p_\xi(y)$$

whose marginal distribution for θ and ψ is the posterior distribution of interest.

3.2.2 Joint exchange algorithm for LERGM and initial state

To constrain ψ to be equal to θ, so that they are no longer distinct, introduces additional computational complexity in (b). In order to put the inferential issue in a form as close as possible to Murray *et al.* (2006), denote the following augmented LERGM likelihood by:

$$p_\theta(x_0, v) = f(x_0, v | \theta) / e^{\zeta(\theta)}$$

where $f(x_0, v | \theta) = \exp\{g(x_0; \theta)\} \vartheta_\rho(R) \prod_{r=1}^{R} T_\theta(v_r; v_{r-1})$. The update of θ would then entail (i) drawing $\xi \sim h(\xi | \theta)$ and (ii) generating $(y, u) \sim p_\xi(y, u)$. Step (ii) now involves generating one initial state y and conditional on this generating a sample path u (no longer constrained to $\prod_{m=1}^{M-1} \mathcal{X}(x(t_{m-1}), x(t_m)))$. While $(u | y, \xi)$ may be sampled using forward simulation, parameters will only be updated when the *entire* sample path is "exchanged". Other algorithms are conceivable but the constraint $\psi = \theta$ means that updating θ involves dealing with the intractability of both $p_\theta(x_0)$ and $p(x(t_1), \ldots, x(t_{M-1}) | \theta, \rho, x_0)$ at the same time. While we want the two distributions to be related through the parameters, we can avoid the complications of the strict requirement that $\psi = \theta$ by introducing a weaker dependency between parameters, compromising between 3.2.1 and 3.2.2. In particular we impose a hierarchal structure, assuming that the parameters ψ and θ follow some distribution, the parameters of which we aim to infer.

3.3 Time heterogeneity

In a previous application of the LERGM to FDI (Koskinen & Lomi, 2013) considerable time-heterogeneity was demonstrated through posterior predictive distributions. By allowing for the model to change piecewise, assuming different parameters $\theta(t)$ for different intervals, goodness-of-fit was considerably improved. In the SAOM framework, Lospinoso *et al.* (2011) have developed a suit of procedures for testing homogeneity of subsets for parameters over time. The approach of Lospinoso *et al.* (2011), in a fully Bayesian framework, would translate to a model selection problem that would involve comparing the marginal likelihoods of many different projections of $\Theta \times [m]$. In principle this could be addressed within a reversible-jump MCMC scheme (Green, 1995) where the moves are constructed to set $\theta_A(t) = \theta_A(t')$ for subsets $A \subset \{1, \ldots, p\}$. This does not depend on the models compared being piecewise constant but when models are no longer indexed by countable index sets this introduces additional complexities. As this is computationally expensive we do not pursue this further here.

Conditional on x_0, we propose to treat $\theta(t)$ as an unobserved time-dependent layer that *a priori* follows some process with density $\phi(\cdot)$ and assuming that the generator $q_{ij}(x | \theta(t), t)$ is conditioned on $\theta(t)$ and that $x(t)$ follow an LERGM conditionally. This affords a very general inference for the time evolution, for example with $\theta(t)$ defined as a multivariate stochastic differential equation. For models ϕ that incorporate elaborate time-dependencies for $\theta(t)$, drawing from the full conditional posterior of $\theta(t)$ given the rest may be cumbersome (and we may not draw on the conditioning out of dependencies that may be used in hidden Markov models).

A convenient alternative may be to let θ change in discrete time and *a priori* assuming an autoregressive form $\theta(t) = A\theta(t-1) + \mu + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma)$, making the

update of $(\theta(t))$ straightforward. This would again mean that the model is piecewise constant but parameters are linked through the model.

3.3.1 Hierarchical model

In the following we will write $\theta(t_0) = \psi$, incorporating the parameters of the ERGM for x_0 in the collection θ of parameters. This symbolically captures the intuition that $\theta(t_0)$ is the parameter vector of the LERGM that started in some state $x(-s)$ for some $s \in \mathbb{R}^+$. We consider here the case when $\theta = (\theta^{(0)}, \ldots, \theta^{(M-1)})$, $\theta^{(m)} = \theta(t_m)$, defines a piece-wise constant process and where $\theta(t_0) = \psi$ is distinct from $\theta(t > t_0)$ but stochastically coupled through $\phi(\theta|\eta)$. In the model for θ, η represents the parameters of interest with prior distribution $\pi(\eta|\gamma)$, for hyper-parameters γ. In particular, we define a hierarchical model similar to Snijders & Koskinen (2012):

$$\pi(\eta|\gamma)$$

$$\times \phi(\theta^{(0)}, \ldots, \theta^{(M-1)}|\eta)$$

$$\times p_{\theta^{(0)}}(x_0) \prod_{m=1}^{M-1} p_{\theta^{(m)}}(x(t_{m-1}), x(t_m))$$

Given a realization θ, updating v is identical to how updating step (a) is carried out with homogenous parameters over time with the only difference being that transition probabilities $T_{\theta(t)}(\cdot|\cdot)$ now are time-specific.

Given a realization v, we may update $\theta(t)$ according to 3.2.1 as all dependence between $\theta(t)$ for different t is captured *a priori* by $\phi(\theta|\eta)$. Updating of $\theta^{(0)} = \psi$ is done as if $\theta^{(0)}$ were distinct from $\theta(t > t_0)$ with $\phi(\theta^{(0)}|\theta(t > t_0), \eta) \propto \phi(\theta^{(0)}, \ldots, \theta^{(M-1)}|\eta)$ assuming the role of prior.

3.3.2 ABC MCMC

As the sample paths $v^{(m)}$ have to be generated using Metropolis, updating steps and mixing may be slow, a considerable amount of time in the algorithm is taken up generating proposals $v^{*(m)}$ and evaluating the complete data likelihood in Equation (3). Approximate Bayesian computation (ABC) relies on the notion that we can draw from $p(D|\theta)\pi(\theta)$ by rejection sampling by proposing values $\theta^j \sim \pi(\theta)$, conditionally on these draw replicate data from the model $D^j \sim p(D|\theta^j)$, and accepting in our sample $\{\theta^j : D^j = D\}$. By relaxing the strict requirement $D^j = D$ and using the criterion $\delta(D^j, D) \leqslant \epsilon$, we get an approximate Bayesian inference for some distance measure δ and tolerance ϵ. If the prior is "far" from the posterior, acceptance rates may be slow, in which case the proposals can be reweighed using some proposal distribution $q(\cdot)$. Here we propose to replace the acceptance probabilities in updating step (b) by the ABC-MCMC equivalent (Bortot *et al.*, 2007; Marjoram *et al.*, 2003):

   i. propose $\theta^{*(m)} \sim q(\cdot|\theta^{(m)})$
   ii. generate $x^{*(m)}$ from the LERGM defined by $p(\cdot|x^{(m-1)}, \theta^{*(m)})$
   iii. if $\delta(x^{*(m)}, x^{(m)}) \leqslant \epsilon_m$, accept $\theta^{*(m)}$ with probability:

$$\min\left(1, \frac{\phi(\theta^{(0)}, \ldots, \theta^{*(m)}, \ldots, \theta^{(M-1)}|\eta)}{\phi(\theta^{(0)}, \ldots, \theta^{(m)}, \ldots, \theta^{(M-1)}|\eta)}\right)$$

Note that this scheme is now conditional on $\theta^{(0)}$, the latter being updated using the approximate exchange algorithm (Caimo & Friel, 2011) with $\phi(\cdot|\eta)$ as a prior. A natural choice for δ is to use the sufficient statistics for the ERGM and the Hamming distance $||x(t_m) - x(t_{m-1})||$ (the corresponding vector of statistics for the curved ERGM (Hunter & Handcock, 2006) would have greater dimension than θ). The rationale for using these statistics is that, while the LERGM is no longer an exponential family model, the ERGM is the limiting distribution of the LERGM. Similarly, using the rationale of the predictive distribution of the statistics under an ERGM (Snijders, 2002) the metric is set to $|w^{\mathrm{T}} z(x^{*(m)})|$, where $z(x^{*(m)})$ is the vector of target statistics and $w = var(z(x^{*(m)}))^{-1}$ is a vector of weights taking into account the variability in statistics. In a pre-tuning phase w is approximated using a Monte Carlo sample of $z(x^{*(m)})$ under some preliminary parameters.

More complicated models for $\theta(t)$ are straightforward to deal with within the ABC framework as long as sampling from these models is easy. For example, the scheme does not rely on the model being piece-wise constant and a continuous-time model for $\theta(t)$ could be accommodated provided the density were analytically tractable up to a normalizing constant.

4 Application to FDI

Building on recent results produced by physicists and economists studying international trade networks (ITN), we analyze the binary architecture of the global FDI network, rather than network flows represented by the magnitude of non-zero cells in the network (Squartini *et al.*, 2011b; De Benedictis & Tajoli, 2011; Head & Mayer, 2014).

An extensive empirical literature in empirical international economics suggests that the so called gravity model of international trade is a useful starting point to model economic relations between countries (Anderson, 2011; Brakman & van Bergeijk, 2010). However, the gravity model of trade is a model for bilateral flows, not network ties.

We adopt the basic version of the gravity model and adapt it to the binary structure of our data. We rely on the gravity model of trade for its recognized empirical success, rather than its theoretical standing, which is still debated in international economics (Anderson, 2011). Our models focus on patterns of local dependence among countries while accounting for gravity factors that are generally known to affect economic relations among countries.

While the application of the gravity model of trade to FDI *flows* is not new (Bevan & Estrin, 2004; Blonigen *et al.*, 2007), the application that we develop to FDI ties deserves additional discussion. FDI flows are determined by streams of investment events involving companies in different countries. We focus on the relation between countries that these event streams establish and, possibly, sustain. In this paper, we choose not to model the underlying event sequences that happen at the level of individual companies. This difference in the level of analysis is important because it illustrates how time-varying streams of events at the micro level (firms) may result in relatively stable relations at more aggregate levels (countries)—stability that justifies a network representation.

The revised gravity model we present specifies the probability of observing a link connecting two countries as proportional to the product of their sizes (the "masses," as measured, for example, by their GDP) and inversely proportional to their geographic distance. The illustration we develop builds directly on Koskinen & Lomi (2013). Our analytical focus is on tie variables. We propose that our approach adds to the existing literature on third-country effects in the gravity model (Feenstra, 2002) in which local dependencies between countries are controlled for generically, but cannot be directly modeled (Anderson, 2011; Anderson & van Wincoop, 2003).

Inspecting the FDI ties across ten years depicted in Figure 1, on the country level, the FDI network demonstrates a great deal of inertia with many stable ties through time (the Jaccard index ranges from 0.096 to 0.24). Inertia is also demonstrated by persistence of local dependencies as evidenced in the models fit in Koskinen & Lomi (2013). Thus it is not only individual ties being entrenched and dyads being reinforced, the same patterns of FDI tend to persist. While there is an inertia, it is also evident from Figure 1 that there is a change. To parse out whether the change in structure over time may be explained by changes in the tie-composition according to one model or whether there is a change to the process itself we need to allow for the model to change over time. A time-heterogeneous model that allows for different parameters for different intervals affords analysis of change of dynamics but it is also desirable to summarize these changes in a consistent manner. It should be obvious that the density varies considerably over time, something that affects the magnitudes of other effects. The hierarchal model is one way of achieving this. Furthermore, although the initial observation is sparse (a density of 0.002) it is vested with some structure. Counts of triadic structures indicate that the dyadic dependencies are not sufficient for explaining the observed clustering (there is one each of the transitive 020 T and mixed 120 U triads, and you only get one or more of these less than 10 times out of 1,000 according to the $\mathscr{U}$|MAN distribution). We also need to start with some initial conditions. While for some types of networks it may be possible to assume an initial clean slate (for example first-year university students), in principle this assumption is unlikely to be equally plausible across all networks analyzed in empirical studies

4.1 Model specification

We use the same model specification as in Koskinen & Lomi (2013), which is a combination of effects that account for local, structural dependencies, as well as "gravity effects". This model specification displayed a good goodness-of-fit for the time-heterogeneous, piece-wise constant model in which $\theta^{(1)}, \ldots, \theta^{(M-1)}$ were estimated. We assume that $g(x; \theta) = \theta^{\mathrm{T}} z(x)$ in Equation (2), where $z(x)$ is a vector of statistics with elements:

1. intercept/density $z_1(x) = \sum_{i,j} x_{ij}$
2. reciprocations $z_2(x) = \sum x_{ij} x_{ji}$
3. alternating out-stars $z_3(x) = \sum_{k=2}^{n-2} (-1)^k s_k^{\mathrm{out}}(x) / \lambda^{k-2}$
4. alternating in-stars $z_4(x) = \sum_{k=2}^{n-2} (-1)^k s_k^{\mathrm{in}}(x) \lambda^{k-2}$
5. alternating triangles $z_5(x) = \lambda \sum_{i,j} x_{ij} [1 - (1 - 1/\lambda)^{L_{2ij}}]$
6. alternating independent 2-paths $z_6(x) = \lambda \sum_{i,j} [1 - (1 - 1/\lambda)^{L_{2ij}}]$
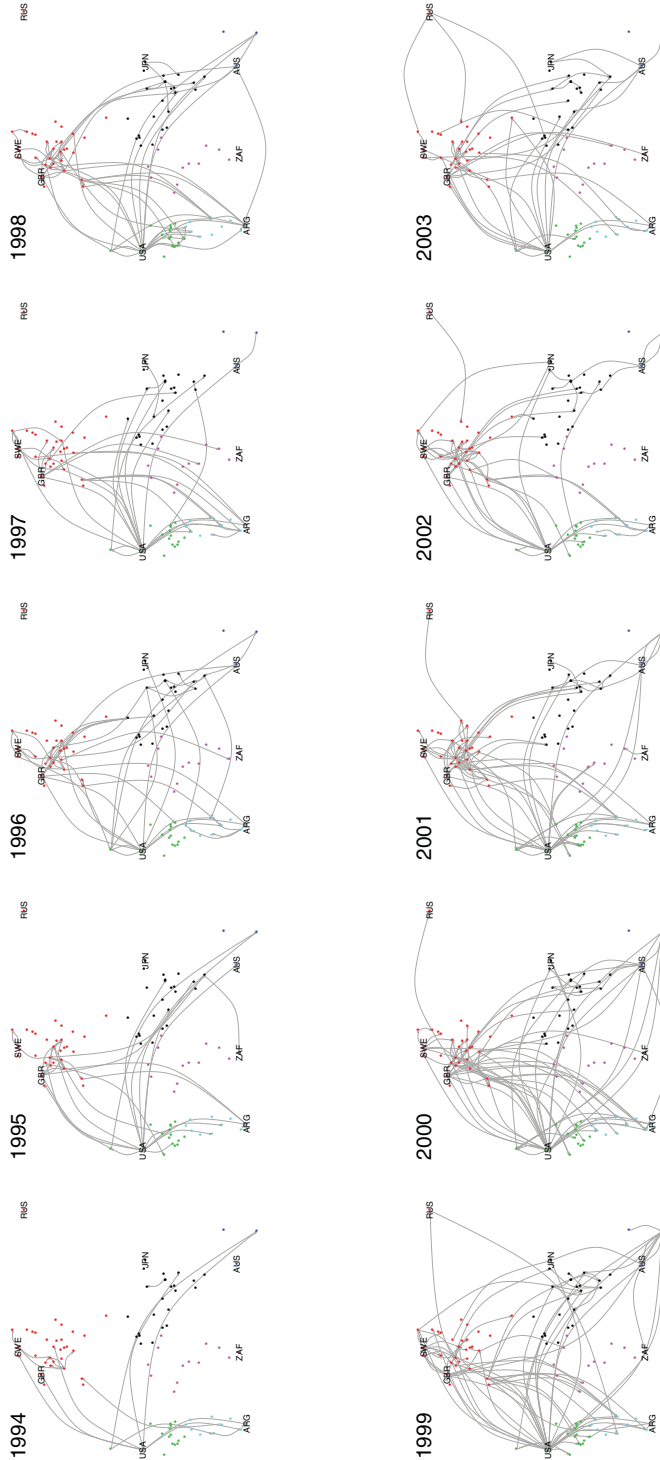
Fig. 1. FDI ties of 97 states over 10 years. Nodes color-coded by continent and laid out in geographical space (Europe magnified) with labels for eight states in the geographical extremes. (Color online)

where L_{2ij} is the number of two-paths between i and j, and $s_k^{\text{out}}(x)$ and $s_k^{\text{in}}(x)$ are the counts of the number of out-k-stars and in-k-stars, respectively. Statistics one through six capture network endogenous dependencies and are derived out of the Markov (Frank & Strauss, 1986) and social circuit dependence assumptions (Snijders *et al.*, 2006). The alternating star effects model the marginal "cost" or "benefit" of sending (receiving) an additional FDI tie given the number of ties the state currently sends (receives). Alternating triangles reflect tendencies towards or against (depending on sign of coefficient) creating ties embedded in transitive structures. Transitive triangles are consistent with local hierarchy and aggregate into clustered regions. Independent two-paths corresponds to indirect ties, where states tend to be tied by many (few) intermediate others.

Covariate-related effects are included as follows:

7. interaction/homophily GDP $z_7(x) = \sum x_{ij} \log V_i \log V_j$
8. sender GDP $z_8(x) = \sum x_{ij} \log V_i$
9. receiver GDP $z_9(x) = \sum x_{ij} \log V_j$
10. distance $z_{10}(x) = \sum x_{ij} \log D_{ij}$

to capture "gravity" effects. Regardless of any underlying rationale based on the gravity model, states are embedded in geographical space and the potential dependencies between ties stemming from spatial effects need to be taken into account. The functional form of the "naive" gravity model also happens to coincide with the spatial interaction function employed in Daraganova *et al.* (2012). An alternative could be to form conditional networks, say conditional on distance deciles as in Abbate *et al.* (2012), allowing further interactions of structural effects and distance. In principle the ERGM framework also allows such interaction of distance and configurations.

We chose a simplistic model $\phi(\theta^{(0)}, \ldots, \theta^{(M-1)}|\eta)$ for the interval-level parameters. Furthermore, we make the simplifying assumption that $\phi(\theta|\eta) = \prod \phi(\theta^{(m)}|\mu, \Sigma)$, where, as in Snijders & Koskinen (2012), $\phi(\cdot|\mu, \Sigma)$ is multivariate normal $N(\mu, \Sigma)$. This allows us a standard normal conjugate model for updating μ and Σ conditional on $\theta^{(0)}, \ldots, \theta^{(M-1)}$. In particular, we assume that *a priori* $\Sigma \sim \text{invwishart}_p(\Lambda_0^{-1}, v_0)$, and conditionally on Σ, $\mu \mid \Sigma \sim N_p(\mu_0, \Sigma/\kappa_0)$. Typically the number of observation points M will be smaller or not much larger than p, meaning that the degrees of freedom v_0 have to be set so as to assure that the posterior (μ, Σ) is proper. For this example we set $v_0 = p + 2$, and $\Lambda_0 = I$. For the rate parameters we assume $\rho^{(m)} \sim \text{gamma}(\alpha, \beta)$ a prior, with $\alpha = 1$ and β large.

A benefit of the Gaussian hierarchical model is that the full conditional posteriors are of standard form. A standard results is that the full conditional distribution of μ given $\theta^{(0)}, \ldots, \theta^{(M-1)}$ and Σ is $N_p(\mu_1, \Sigma_1/(\kappa_0 + M))$, where

$$\mu_1 = \frac{M}{\kappa_0 + M}\bar{\theta} + \frac{\kappa_0}{\kappa_0 + M}$$

and $\bar{\theta} = 1/M \sum_m \theta^{(m)}$. The full conditional posterior of Σ is $\text{invwishart}_p(\Lambda_1^{-1}, v_0 + M)$, where

$$\Lambda_1 = \Lambda_0 + \mathbf{Q} + \frac{\kappa_0 M}{\kappa_0 + M}(\bar{\theta} - \mu_0)(\bar{\theta} - \mu_0)^{\text{T}}$$

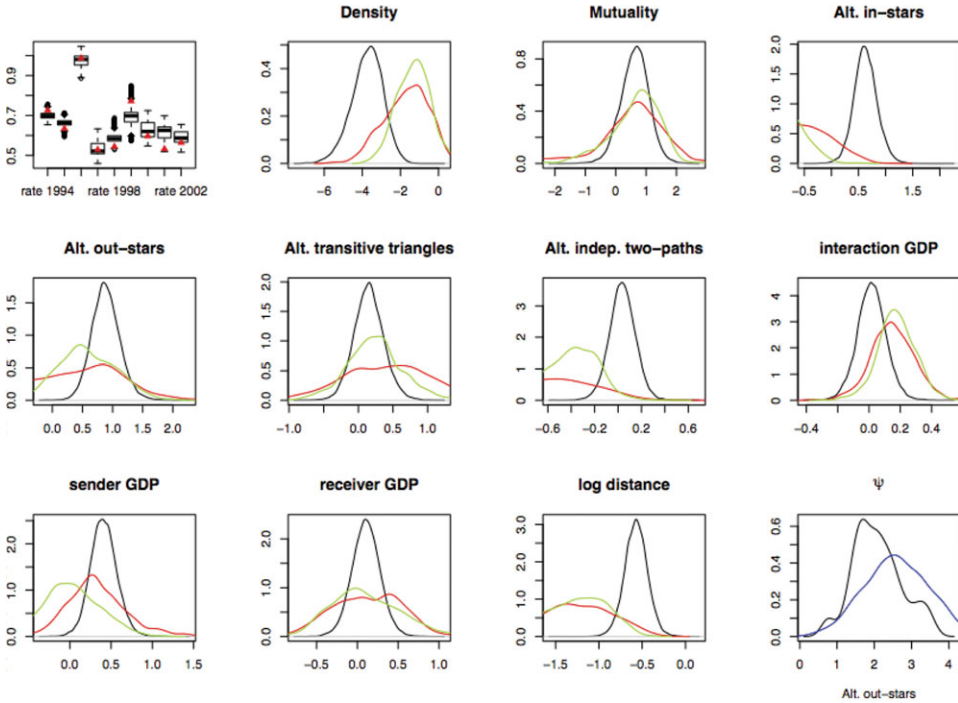and $\mathbf{Q} = \sum_m (\theta^{(m)} - \bar{\theta})(\theta^{(m)} - \bar{\theta})^{\text{T}}$.

Fig. 2. Posterior distributions for $\rho^{(m)}$ (top left), μ, and $\theta^{(0)}$. MCMC posteriors for μ (black), ABC with tolerance $\epsilon = 7$ (red), and $\epsilon = 4$ (green). Bottom right provides the density estimate for $\theta^{(0)}$ from hierarchical model (black) and fitted separately (blue). (Color online)

The estimation procedure was implemented in Matlab (code available upon request) and a partial implementation is forthcoming as a routine in R that draws on the package Bergm (Caimo & Friel, 2014).

4.2 Results

The overall substantive results of the analysis are similar to that Koskinen & Lomi (2013). However, their analysis was limited to the choice of either a time-homogenous model, which was shown to have a poor fit, or a time-heterogeneous model, which demonstrated a good fit but only offered estimates for each interval separately. The main target of inference here is the posterior distribution for μ given in Figure 2, which is a joint inference for the entire period that takes into account the variation across intervals as well as the model for the initial observation. The posteriors for μ may be considered a consistent way of summarizing the dynamics over time, a parametric pooling of the interval-level estimates. While the top-level parameters μ are of primary interest we may also investigate the posterior predictive distributions for $\theta^{(0)}, \ldots, \theta^{(M-1)}$

$$\int \pi(\theta^{(0)}, \ldots, \theta^{(M-1)}, \mu, \Sigma | x_0, x(t_1), \ldots, x(t_{M-1})) d\mu d\Sigma$$
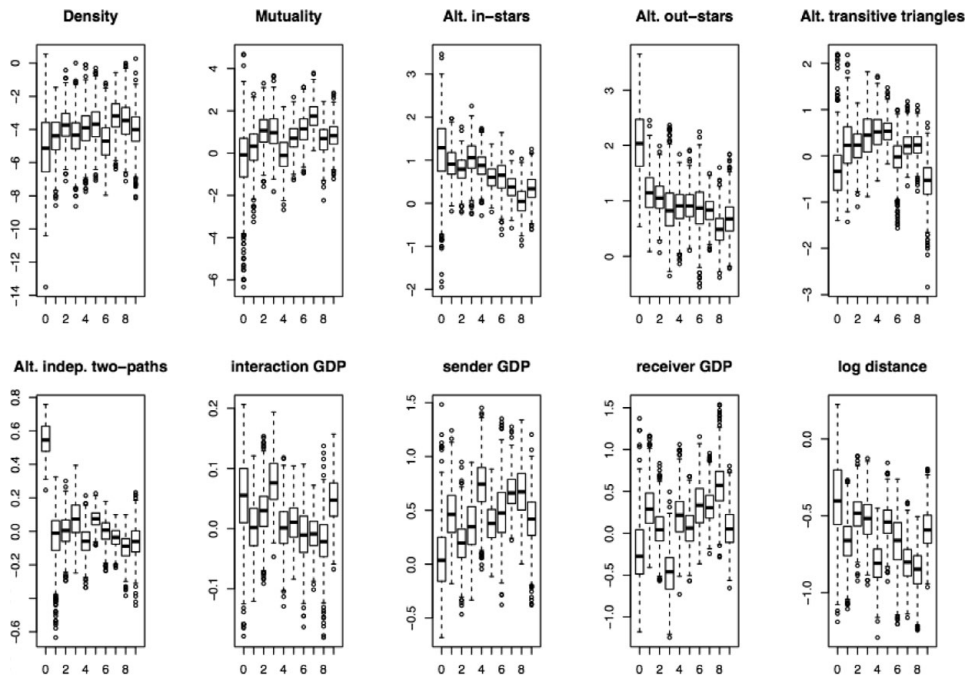
Fig. 3. Posterior predictive distributions for $\theta^{(0)}, \ldots, \theta^{(M-1)}$ by coordinate and year.

These are the marginal predictive distributions for the interval-level parameters given data plotted in Figure 3[1].

In summary, the dependence between FDI ties stemming from endogeneity induced by the sender of the tie is evidenced in the alternating out-degree effect. This reflects that there are costs and benefits associated with establishing an additional tie $i \rightarrow h$ for i, that depends on present commitments. This is also concordant with a "rich-get-richer" (or positive feedback) effect possibly produced by a decline in the costs (and risks) of entering a foreign market after the first entries. There is no evidence of transitivity in the FDI ties despite the significant triad census profile for the cross-sections reported in Koskinen & Lomi (2013). We would perhaps have expected some evidence of local hierarchy above the global hierarchy evidenced by the degree-based effects. A possible explanation is that the gravity components account for this.

What we learn from comparing the posterior for μ with the posterior for θ in the homogenous model ($\theta^{(m)} = \theta$, for all m; Figure 6 of Koskinen & Lomi, 2013) is that the extra layer of parameters, allowing $\theta^{(m)}$ to vary about a mean μ, comes with additional uncertainty. This is evidenced in wider posteriors for μ than for the time-homogeneous θ. Only some of this additional uncertainty is contributed by the initial observation. For example, the point estimate of the mutuality parameter in the homogenous model is 0.88 (with standard deviation 0.25) with a posterior

---

[1] A error was detected in the data for 2000 used by Koskinen & Lomi (2013). We have left this uncorrected for comparability. Correcting the data has the effect of increasing the marginal estimates (with constant prior) for the interval 1999–2000 of alternating out-stars and triangles from 1.1 and −0.21 to 2.2 and 0.27, respectively, and attenuating the GDP effects, but leaving everything else largely unchanged.

probability of being positive of nearly unity. The corresponding parameter in the hierarchal model is 0.64 (with standard deviation 0.46) with only a 0.92 probability of being positive. This is partly a result of the evidence for reciprocity provided by x_0 (reflected in $\theta^{(0)}$ in Figure 3) pulling μ towards zero. The initial observation has a different effect on the parameter for alternating out-stars (the posterior mean for the hierarchal model is 0.89 compared to 0.84 for the time-homogenous model). The time-homogenous model where $\theta^{(m)}$ is forced to be equal to θ for each interval m is miss-specified and gives artificial precision through pooling information across intervals arbitrarily.

There are considerable differences in rates of change over the period. The rates for the intervals 1994–95 and 1995–96 are around 0.7 but the rate for 1996–97 is close to 1.0 (top left panel of Figure 2). This closely resembles the results of Koskinen & Lomi (2013) both for their time-homogenous and time-heterogeneous models (recall the difference in parameterization for ρ noted in the description of updating step c above). The mixing for the rate parameters in the chain is poor. Conditional on a current path-length $r^{(m)}$, $\rho^{(m)}$ is drawn using a Gibbs updating step, in one block. This is fast but moves between different length chains, i.e. increasing or decreasing $r^{(m)}$, is slow. The considerable variability in path length and consequently in ρ is related to the fact that rates for the holding times are constant. Increasing the rate ρ implies that an increased number of tie-variables will be updated. The number of tie-variables that *have to change* is however extremely small in proportion to the total number of tie-variables. However, the conditional distributions for $\theta^{(m)}$ appear to be more or less independent of the rate $\rho^{(m)}$ (results not presented here).

For the ABC-MCMC, we fixed the rate parameters at the values indicated by red triangles in the top left panel of Figure 2. As expected the uncertainty in the ABC posterior is reduced by decreasing the tolerance ϵ. Decreasing the tolerance accentuates the difference between the ABC posterior and the "true" posterior distribution. This is particularly obvious for alternating in-stars, where we see a shift in the distribution to the left. As noted in Section 3.3.2, the ABC posterior would converge to the true posterior as the tolerance tends to zero had the LERGM been an exponential family distribution with sufficient statistics $z(\cdot)$. The discrepancies between the MCMC posteriors and the ABC MCMC posteriors illustrate the fact that the dependency through time from the conditioning in the LERGM renders it different from its limiting ERGM. Notable deviations are the posteriors for the structural effects in-stars and two-paths. This may also affect the interaction and sender GDP effects.

The posterior predictive distributions for $\theta^{(m)}$ in Figure 3, obtained in the process of performing inference for μ, suggest that the parameters are serially autocorrelation. This is not surprising as the simplistic model we have used does not account for time-dependence in the parameters. The predictive distributions for $\theta^{(0)}$ are generally more spread out than for $m > 0$ reflecting the fact that x_0 is very sparse and the distribution of ψ not using the hierarchical structure is uncertain. The bottom right hand panel of Figure 2, demonstrates the effect on the distribution of for example the alternating out-star parameter by imposing the model $\phi(\cdot|\mu, \Sigma)$. The inference for x_0 thus borrows power from the subsequent process and its posterior is pulled towards the mean μ although the bottom right panel of Figure 2 illustrates that we can also fit a model for x_0 both separately and jointly.

The posterior predictive distributions in Figure 3 demonstrate the added value of the model for the initial state. With one exception, there is a remarkably good correspondence between $\theta^{(0)}$ and the subsequent $\theta^{(m)}$'s for $m > 0$. For density and the covariate dependent parameters the distribution for $\theta^{(0)}$ is roughly in the range of the other parameters (allowing for the added uncertainty due to the sparseness of the first network). For alternating in- and out-stars there is a smooth transition from t_0 and onwards following a clear downward trend. If we were to interpret this as evidence of decreasing in- and out-degree centralization over time, the additional modeling of the initial state accentuates these trends. Looking at the networks in Figure 1, we see that in 1994 there are two big players, the USA and France, that account for most ties, but that the FDI networks become gradually less heterogeneous as the electricity market matures.

Similarly, for alternating triangles the distribution for x_0 makes the pattern of change to transitivity over time more visible. The tendency towards transitivity appears to increase over time until 1999 after which the effect decreases, becoming negative for the last interval. The agreement in $\theta^{(0)}$ with $\theta^{(m)}$ for $m > 0$ in terms of magnitude, as well as trends, demonstrates that it is meaningful to assume an ERGM for the initial state even if the ERGM for the initial x_0 has to account for the potentially accumulated time-heterogeneity of several years, whereas the parameters $\theta^{(m)}$ ($m > 0$) only need to describe the dynamics of a one-year interval. The notable exception to this is the alternating independent two-path parameter. The "pooled" estimate (μ) is very close to zero and all predictive distributions are centered on the origin with the exception of $\theta^{(0)}$ for which there is strong evidence that it is positive. This may be interpreted as the initial network being more connected than we would expect, everything else equal. Two components, one centered on the USA and the other on France, contribute many such two-paths. The latter one, consisting of France, Belgium, Italy, the Netherlands, Sweden and Finland, is tied together by eight ties with only the tie between France and Germany being reciprocated.

Figure 4 illustrates the benefit of the added layer μ and Σ compared to the modeling formulation in Koskinen & Lomi (2013). The mixing for the interval-level parameter associated to alternating in-stars is bad in the sense that there is a lot of serial autocorrelation (possibly due to the proposals of the paths v). However, the mixing of the chain in the upper level for global parameter μ is excellent. Considering that the posteriors for the average across intervals (μ) is very close to the posteriors ($\theta^{(m)} = \theta$, for all m) in the time-homogenous model, with the same substantive interpretation of effects, but with a much greater efficiency, the hierarchical model is preferable even if it does not take into proper account associations between parameters over time.

5 Conclusions

We have demonstrated an elaboration of the longitudinal ERGM that allows for estimating effects for the initial condition as well as a parametric form for the change over time by exploiting a simple and flexible hierarchical representation of network dynamics. Bayesian techniques transfer seamlessly across different domains and models, providing a compact, unified approach to analysis. The analysis of FDI relations between countries observed over a decade produced evidence of network
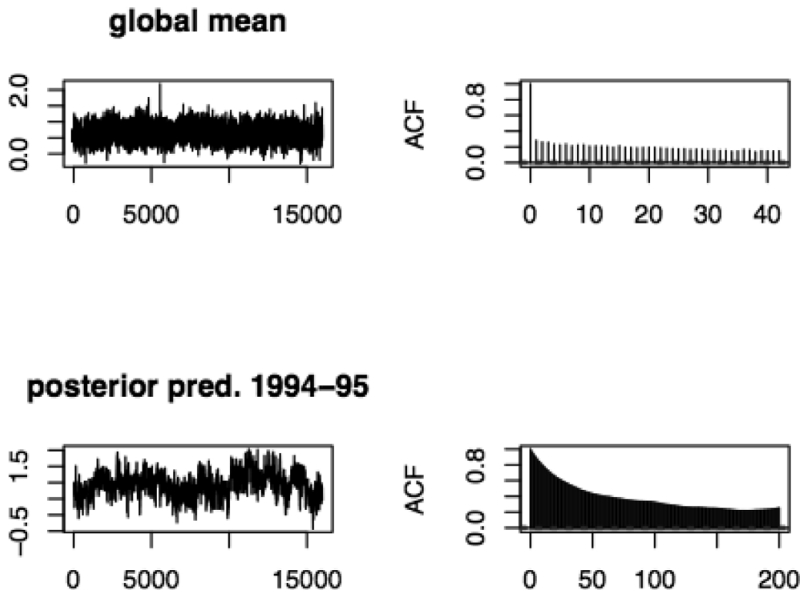
Fig. 4. Comparison of posterior draws for global mean parameter μ and posterior predictive $\theta^{(m)}$ ($m = 1$) for interval 1996–97 (alternating in-stars).

centralization (both in in- as well as out-degree), of strong spatial effects, and of a strong country size effect on sending FDI ties. These conclusions take into account the variation across years and the interdependence with other effects such as, for example, varying densities. The model captures well the co-evolution of firms target selection strategies behind observed FDI decisions, and the global network of relations between countries in the specific industry we have examined.

In the new modeling framework we have proposed, inference relies on recent advanced Bayesian computational methods. More specifically. We adopted an approximate exchange sampler based on the algorithm of Caimo & Friel (2011) to carry out estimation for initial conditions. The ABC-MCMC likelihood-free approach affords an easier and less computationally intensive estimation of the hierarchical model. The likelihood-free nature of ABC-MCMC also offers additional flexibility in modeling complex time-heterogeneity. The ABC-MCMC approach relies on intuitions very similar to non-Bayesian stochastic approximation techniques frequently used for non-longitudinal ERGMs. Here it offers considerable gains is estimation speed, but appears to induce some systematic errors that may be due to the LERGM not being an exponential family distribution. The ABC-MCMC is however one of the more simplistic likelihood-free approaches and elaborations that have proved successful for similar problems may improve on the accuracy of the results. These elaborations also provide natural alternatives for the exchange sampler such as particle MCMC (Andrieu *et al.*, 2010), something that would open up for modeling initial states for longitudinal models where no natural first observation model suggests itself (for example the SAOM).

The parametric framework for handling changing dynamics over time is very general and has the benefit of improving inference considerably. As is well-known in machine learning, adding layers deals effectively with complicated dependencies

and heterogeneous information. A computational fringe benefit in the context of our analytical framework is that the inference for μ "smooths" the inference for the lower-level parameters $\theta^{(m)}$.

The main benefit of the extension of the model is that it provides us with a procedure for a joint analysis of data for a process that is obviously heterogeneous through time. The hierarchal framework allows further parameterizing the change in the process, something that the time-homogenous and time-heterogeneous models of Koskinen & Lomi (2013) were unable to provide. In particular, the marginal analysis of the interval-level parameters suggested that the model may be further improved by modeling the decreasing centralization and cyclic transitivity parametrically. A straightforward elaboration of the hierarchal model used here would be to explicitly to let the dynamics depend on time as $\theta = \mu + \beta \mathbf{W} + \epsilon$, for $\epsilon \sim N_M(\mathbf{0}, \Sigma)$, and where $\mathbf{W} = (W_0, \ldots, W_{M-1})$ for "network-level" predictors W_m. For a particular time-period m, W_m may incorporate $(t_m - \bar{t})$ and $(t_m - \bar{t})^2$ to capture the dependence on time. As these covariates are defined on the network-level they may in addition include world-level covariates. Adding more waves and refining the functional form of the time-heterogeneity will enable us to elaborate the functional form of these changes. An alternative to a smooth parametric change to the parameters could be inferring change-points in the process.

For the particular example of FDI, the first observation at 1994 was very sparse and did not affect the overall inference dramatically. However, the marginal analysis for the initial observation (the posterior predictive distribution) suggests that an ERGM for the initial state contributes information that is consistent—and may be interpreted alongside with the inference for the subsequent LERGM. In particular, the additional estimates highlight the pattern of the change to the dynamic process over time. Many empirical data sets may have non-trivial first observations that might affect significantly the longitudinal analysis. For example, a longitudinal model that fails to account for an initial state characterized by high levels of homophily, may erroneously suggest the conclusion that homophily is not a significant data-generating mechanism for the specific sample that is being analyzed. As an example, Igarashi (2013) discusses whether the lack of gender effects in a LERGM analysis of two waves of face-to-face interaction may be due to gender-biases primarily affecting the early stages of interaction. Simultaneously fitting an ERGM and a LERGM may directly answer that question.

References

Abbate, A., De Benedictis, L., Fagiolo, G., & Tajoli, L. (2012). The international trade network in space and time. *LEM Working Paper Series, Institute of Economics Scuola Superiore Sant'Anna*. Available at SSRN: http://ssrn.com/abstract=2160377

Anderson J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, **69**, 106–116.

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, **3**, 133–160.

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, **93**, 170–192.

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B* (Statistical Methodology), **72.3**, 269–342.

Bergstrand J. H. (1985). The gravity equation in international trade: Come microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, **67**, 474–481.

Bevan, A. A., & Estrin, S. (2004). The determinants of foreign direct investment into European transition economies. *Journal of Comparative Economics*, **32**, 775–787.

Blonigen, B. A., Davies, R. B., Waddell, G. R., & Naughton, H. T. (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review*, **51**, 1303–1325.

Bortot, P., Coles, S. G., & Sisson, S. A. (2007). Inference for stereological extremes. *Journal of the American Statistical Association*, **102**, 84–92.

Brakman, S., & van Bergeijk, P. (2010). *The gravity model in international trade: Advances and applications*. Cambridge: Cambridge University Press.

Caimo, A., & Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, **33**, 41–55.

Caimo, A., & Friel, N. (2014). Bergm: Bayesian exponential random graphs in R. *Journal of Statistical Software* **61**(2).

Chakrabarti A. (2003). A theory of the spatial distribution of foreign direct investment. *International Review of Economics & Finance*, **12**, 149–169.

De Benedictis, L., & Tajoli, L. (2011). The world trade network. *The World Economy*, **34**, 1417–1454.

Daraganova, G., Pattison, P., Koskinen, J., Mitchell, B., Bill, A., Watts, M., & Baum, S. (2012). Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. *Social Networks*, **34**, 6–17.

Dueñas, M., & Fagiolo, G. (2013). Modeling the international-trade network: A gravity approach. *Journal of Economic Interaction and Cooridination*, **8**, 155–178.

Fagiolo, G., Schiavo, S., & Reyes, J. (2009). World-trade web: Topological properties, dynamics, and evolution. *Physical Review E*, **79**, 036115.

Feenstra, R. C. (2002). Border effects and the gravity equation: Consistent methods for estimation. *Scottish Journal of Political Economy*, **49**, 491–506.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, **81**, 832–842.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In Gopinath, Helpman, & Rogoff (Eds.), *Handbook of international economics,* volume 4. Amsterdam: Elsevier.

Holland, P. W., & Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology*, **5**, 5–20.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, **76**, 33–65.

Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, **15**, 565–583.

Igarashi, T. (2013). Longitudinal changes in face-to-face and text message-mediated friendship networks. In D. Lusher, J. H. Koskinen, & G. F. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods and applications* (pp. 248–259). New York: Cambridge University Press.

Koskinen, J. H., & Lomi, A. (2013). The local structure of globalization: The network dynamics of foreign direct investments in the international electricity industry. *Journal of Statistical Physics*, **151**, 523–548.

Koskinen, J. H., & Snijders, T. A. B. (2007). Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference*, **137**, 3930–3938.

Lospinoso, J. A., Schweinberger, M., Snijders, T. A. B., & Ripley, R. M. (2011). Assessing and accounting for time heterogeneity in stochastic actor oriented models. *Advances in Data Analysis and Computation*, **5**, 147–176.

Marjoram, P., Molitor, J., Plagnol, V., & Tavare, S. (2003). *Markov chain Monte Carlo without likelihoods Proceedings of the National Academy of Sciences of the United States*, **100**, 324–328.

Murray, I., Ghahramani, Z., & MacKay, D. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*. AUAI Press, Arlington, Virginia.

Pattison, P., & Robins, G. L. (2002). Neighbourhood-based models for social networks. *Sociological Methodology*, **32**, 301–337.

Redding, S. J. (2011). Theories of heterogeneous firms and trade. *Annual Review of Economics*, **3**, 77–105.

Robins, G. L., & Pattison, P. E. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, **25**, 5–41.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. In M. E. Sobel, & M. P. Becker (Eds.), *Sociological Methodology* (pp. 361–395). London: Blackwell.

Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, **3**(2).

Snijders, T. A. B. (2006). Statistical methods for network dynamics. In S. R. Luchini (Ed.), *XLIII Scientific Meeting, Italian Statistical Society* (pp. 281–296). Padova: CLEUP.

Snijders, T. A. B., & Koskinen, J. (2013) Longitudinal models. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods and applications* (pp. 130–140). New York: Cambridge University Press.

Snijders, T. A. B., & Koskinen, J. (2012). Multilevel longitudinal analysis of social networks. Paper presented at the *8th UKSNA Conference*, Bristol, June 28–30. `http://www.stats.ox.ac.uk/~snijders/siena/siena_articles.htm`

Snijders, T. A. B., Koskinen, J. H., & Schweinberger, M. (2010). Maximum likelihood estimation for social network dynamics. *Annals of Applied Statistics*, **4**, 567–588.

Snijders, T. A. B., Pattison, P., Robins, G., & Handcock, M. (2006) New specifications for exponential random graph models. *Sociological Methodology*, **36**, 99–153.

Squartini, T., Fagiolo, G., & Garlaschelli, D. (2011a). Randomizing world trade. I. A binary network analysis. *Physical Review E*, **84**, 046117.

Squartini, T., Fagiolo, G., & Garlaschelli, D. (2011b). Randomizing world trade. II. A weighted network analysis. *Physical Review E*, **84**, 046118.

Wasserman, S. (1980). Analyzing social networks as stochastic processes. *Journal of the American Statistical Association*, **75**, 280–294.

Wasserman, S., & Pattison, P. E. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, **61**, 401–425.